

jc914 U.S. PRO  
09/747160  
12/22/00

APPENDICES I, II AND III  
FOR  
APPLICATION  
FOR  
UNITED STATES LETTERS PATENT

TITLE: COLLECTING AND ANALYZING SURVEY DATA  
APPLICANTS: LANCE PARKER, FERNANDO ALVAREZ and MICHAEL  
H. COEN

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL227256583US

I hereby certify under 37 CFR §1.10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231

December 22, 2000

Date of Deposit

Victor Mahoney

Signature

Victor Mahoney  
Typed or Printed Name of Person Signing Certificate

Appendix I

Semantic Tagging ("tagging") is a process of  
formatting individual questions and responses in a survey  
5 in a formal, machine-readable knowledge representation  
language (KRL) to enable automated analysis of data  
obtained via that survey. The semantic tags (or simply  
"tags") indicate the meaning of a response to a question in  
a particular way. The tags are created by a survey author  
10 (either a person, a computer program, or a combination  
thereof) and allow engine 30 to understand both a question  
and a response to that question.

Tags indicate what the gathered information actually  
represents and allow the engine 30 to process data  
15 autonomously. In particular, the tags allow the data  
collected by a survey to be directly processed by an expert  
(i.e., rule-based) or logic programming (e.g., Prolog-  
based) system in engine 30 without requiring direct human  
intervention to interpret, categorize, summarize, etc.,  
20 survey responses. User responses are asserted as facts  
within an expert system (e.g., within engine 30), where  
each fact is automatically derived from the tag associated

with each question.

Tags represent the information gathered by a particular question, but are not tied to the precise wording of that question. Thus, it is possible for a wide  
5 range of natural language questions to have identical semantic tags. The KRL here has a partial ontology for describing survey questions and responses. It is intended to be descriptive and functional and thereby capture the vast majority of questions on marketing surveys.

10 In this embodiment, surveys are comprised of three types of questions: behavioral, attitudinal, and demographic. Each of these question types has a corresponding unique type of tag which, as noted above, includes question and response fields. Examples of the  
15 question fields of these tags are set forth below. In the tags, the following conventions apply:

- (1) "|" represents a logical OR operation
- (2) plain text represents constants or list headers  
in an s-expression
- 20 (3) **bold print** represents keyword arguments
- (4) *italics* represent a member of a named set
- (5) <brackets> surround optional items

(6)  $\{NAME\}$  refers to a variable

## 1.0 Behavioral Questions

The question field tag template for behavioral  
5 questions is as follows.

```
(tag      (type behavioral)
  (time
    (tense past | present | future)
10    (startDate  date)
      <(endDate date)>
  )
  (activity
    (action act | act AND act | act OR act)
15    (queryRegarding quality)
      (object product)
        <(subject demographic)>
        <(indirectObject demographic)>
        <(verb string)>
20    <(variable string)>
      )
    <(questionID string)>
    (response ...))
```

25  $act \in \{\text{Use, Do, Purchase, Replace, License, Own, Sell,}$   
Exchange, Recommend, Repair, Visit, Contact, Complain, and  
similar expressions}

$quality \in \{\text{Frequency, Length, Existence, Source,}$   
30 Intention, Purpose, Completion, Difficulty, and similar  
expressions}

*string* is a "quotation delimited" string of characters.

*product* is a set of products and/or services offered by a particular client and is industry specific. It is enumerated when the expert system is first installed for a client and subsequently can be modified to reflect the evolution of the client's product line or the client's industry as a whole. Elements of the *product* set have ad hoc internal structure representing both the client's identity and an item's position in the client's overall hierarchy of product/service offerings. By way of example, an IBM laptop computer is represented by

"IBM/product/computer/laptop."

The *demographic* set is defined in the section of templates for demographic questions set forth below.

The *response* field that corresponds to the above question field is specified below.

Each individual question in a survey has a tag that adheres to the above template, but need not assign optional fields. For example, consider the following behavioral questions and their associated tags immediately following  
5 the questions.

(1) "Have you used an IBM laptop computer in the past 3 years?"

10 (tag (type **behavioral**)  
(time  
(tense **past**)  
(startDate (\${CURRENT\_DATE} - 3 YEARS))  
(endDate \${CURRENT\_DATE}))  
15 (activity  
(action **Use**)  
(queryRegarding **Existence**)  
(object "IBM/product/computer/laptop")  
)  
20 (response  
(type **YorN**)))

(2) "How often do you replace your server?"

25 (tag (type **behavioral**)  
(time  
(tense **present**)  
(startDate \${CURRENT\_DATE}))  
)  
30 (activity  
(action **Purchase**)  
(queryRegarding **Frequency**))

```
        (object "any/product/server")
      )
    (response
      (type Selection)
5      (selections (choseOne 0 3 6 9 12 18))
      (primitiveInterval Month)))
```

(3) "What brand of laptop computer do you use now?"

```
10  (tag      (type behavioral)
      (time
        (tense present)
        (startDate  ${CURRENT_DATE})
      )
15    (activity
        (action Use)
        (queryRegarding Source)
        (object "current/product/computer/laptop")
        (variable "CURRENT_BRAND")
20    )
    (response
      (type MenuSelection)
      (selections (onlyOne "IBM" "Compaq" "NEC" "Gateway"
25      "Dell" "Sony" "HP"))
      (setVariable "CURRENT_BRAND"))))
```

## 2.0 Behavioral Questions

30 The question field tag template for attitudinal questions is as follows.

```
(tag      (type attitudinal)
      (time
        (tense past | present | future)
35      (startDate  date))
```

```

        <(endDate date)>
    )
    (attitude
      (belief belief )
5      (queryRegarding beliefQuality)
      <(statement string)>
      <(subject demographic)>
      (object reference)
      (attribute feature)
10      <(contrast reference)>
      <(variable string)>
    )
    <(questionID string)>
    (response ...))
15
```

*belief* ∈ {Satisfaction, Perception, Preference, Agreement, Modification, Plausibility, Reason, and similar expressions}

20 *beliefQuality* ∈ {Degree, Correlation, Absolute, Ranking, Specification, Elaboration, and similar expressions}

*feature* is a set of features relevant to a particular client's product and/or service offerings. Although many features are industry specific, many such as *reliability* are fairly universal. The *feature* set is enumerated when the expert system is first installed for a client and can be subsequently modified to reflect the evolution of the client's product line or industry as a whole.



The *response* field that corresponds to the above question field is specified below.

5           It is noted that, for questions with matrix scales, which is a common occurrence in attitudinal questions, each row of the matrix has a separate, unique tag.

10           Consider the following attitudinal questions and their associated tags, which immediately follow the questions.

(1) "Rate your overall satisfaction with the performance of the laptop computer you are currently using."

15   (tag           (type **attitudinal**)  
          (time  
            (tense **present**)  
            (startDate   \${CURRENT\_DATE}))  
          )  
20    (attitude  
          (belief **Satisfaction** )  
          (queryRegarding **Degree**)  
          (object "current/product/laptop/computer")  
          (attribute **Performance**))  
25    (response  
          (type **HorizontalLikert**)  
          (askingAbout **Satisfaction**)  
          (selections   (low 0)  
                        (high 5)  
30       (interval 1)))

(2) "If you could make one change to your current laptop computer, what would it be?"

```
5  (tag      (type attitudinal)
      (time
        (tense present)
        (startDate  ${CURRENT_DATE}))
    )
      (attitude
10      (belief Modification )
        (queryRegarding Specification)
        (object "current/product/laptop/computer"))
      (response
15      (type ListSelection)
        (selections (onlyOne ${FEATURES})))
```

(3) "Do you agree with the sentiment that laptop computers will someday replace desktop computers?"

```
20  (tag      (type attitudinal)
      (time
        (tense present)
        (startDate  ${CURRENT_DATE}))
    )
      (attitude
25      (belief Agreement )
        (queryRegarding Absolute)
        (object "any/product/laptop/computer")
        (statement "Laptop computers will someday replace
30 desktop computers."))
      (response
        (type YorNorDontKnow)
        (askingAbout Agreement))
```

35 (4) "Do you have any additional comments to add?"

```
(tag      (type attitudinal)
  (time
    (tense present)
5      (startDate  ${CURRENT_DATE})
    )
    (attitude
      (belief Perception )
      (queryRegarding Elaboration)
10     (object "any/product/laptop/computer")
    )
    (response
      (type FreeResponse)
      (noLines 3)
15     (width 40)
    )
  )
```

### 20 3.0 Demographic Questions

The question field tag template for demographic questions is as follows:

```
(tag      (type demographic)
25  (time
    (tense past | present | future)
    (startDate  date)
    <(endDate date)>
  )
30  (description
    <(gender)>
    <(age)>
    <(ageRange)>
    <(haveChildren)>
35  <(numberChildren)>
    <(childAgeByRange)>
    <(maritalStatus)>
  )
)
```

```

5      <(employment)>
      <(education)>
      <(income)>
      <(address)>
      <(email)>
      <(name)>
      <(phoneNumber)>
      <(faxNumber)>
10     <(city)>
      <(state)>
      <(zipCode)>
      <(publicationsRead)>
      <(groupMembership)>
      <(hobbies)>
15     <(mediaOutlets)>
      <(other string)>
      <(qualifier length | prefer | like | dislike
| know | dontKnow)>
      )
20     <(questionID number)>
      (response ...)
```

The *response* field for the above question field is specified below.

25

By way of example, consider following the demographic questions and their associated tags immediately following the questions.

30 (1) "What is your gender?"

```
(tag      (type demographic)
  (time
    (tense present)
```

```
        (startDate    ${CURRENT_DATE}))
      (description
        (gender))
      (response
5       (type Selection)
        (selections (onlyOne "Male" "Female"))))
```

(2) "What is your email address?"

```
10  (tag      (type demographic)
      (time
        (tense present)
        (startDate    ${CURRENT_DATE}))
      (description
15       (email))
      (response
        (type FreeResponse)
        (noLines 1)
        (width 30)
20     )
    )
```

(3) "How long have you lived at your present address?"

```
25  (tag      (type demographic)
      (time
        (tense present)
        (startDate    ${CURRENT_DATE}))
      (description
30       (address)
        (qualifier length)
      )
      (response
35       (type MenuSelection)
        (low 0)
        (high 20+)
```

```
        (primitiveInterval Year)
    )
)
```

5

#### 4.0 Response Field Template

Questions in surveys can have a variety of different scales for allowing the respondent (i.e., the one taking the survey) to select an answer. The response field of a tag specifies, for each question in the survey, both the  
10 general scale-type that the response field uses and how to instantiate that scale to obtain a valid range of answers.

The response field also contains placeholders for the respondent's actual answers and individual (perhaps  
15 anonymous) identifier(s). Each completed survey for some respondent leads to all of the tags associated with that survey being asserted as facts in the expert system, with all of the placeholders appropriately filled in by the respondents' answers. For expert systems, such as CLIPS  
20 (C-Language Integrated Product System) that do not support nested structures within facts, the actual data representation is a flattened version of the one shown below.

A representative template for the response field is as follows.

```
(response
5   (type scale)
    <(askingAbout questionTopic)>
    <(prompt string)>
    <(low number)>
    <(high number)>
10   <(interval number)>
    <(scaleLength number)>
    <(primitiveInterval time | distance | temperature)>
    <(selections (onlyOne | anyOf string+)
                <(upto number)>
15                <(atLeast number)>
    )>
    <(width number)>
    <(noLines number)>

20   (userSelectionRaw string | number)
    (userSelection string)
    (userSelectionType string)
    (userID number)
    (userIDinternal number)
25   (userIDconfidential string)
    (clientID string))
```

*scale* ∈ {Likert, Selection, MenuSelection, YorN,  
YorNorDontKnow, FreeResponse, HorizontalLikert}

30

The *askingAbout* field can be set to have the expert system automatically generate the prompt for selecting an answer.

*questionTopic* ∈ {Preference, Sentiment, Belief, Frequency,

Comment, and similar expressions}

## 5.0 Fact Instantiation from Tags

5       Tags allow the data collected by a survey to be  
directly processed by an expert (i.e., rule-based) or logic  
programming (e.g., Prolog-based) system (engine 30) without  
requiring direct human intervention to interpret,  
categorize, summarize, etc., survey responses. User  
10    responses are asserted as facts within the expert system,  
where each fact is automatically derived by parsing the  
relevant information from a corresponding tag associated  
with each question.

It is noted that that additional information regarding  
15    each user is simultaneously instantiated in separate facts  
within the expert system. This includes for example, the  
site where the respondent was surveyed, the time of day the  
survey was taken, and the like.

By way of example, consider the question:

20

"How often do you speak with your salesman?",

with associated tag:



```
(tag      (type behavioral)
  (time
    (tense current)
5    (startDate ${CURRENT_DATE}))
  )
  (activity
    (action Contact)
    (queryRegarding Frequency)
10    (indirectObject "NEC/person/salesman")
    (object "NEC/product/PBX/NEAX2000"))
  (response
    (type Selection)
    (askingAbout Frequency)
15    (selections (onlyOne 0 3 6 9 12 18))
    (primitiveInterval Month)))
```

If a respondent answering this question selects "3",  
as in, "I speak with my salesman every 3 months", the  
20 expert system will automatically assert a fact  
corresponding to the tag in the expert system, with  
additional fields representing the user's selection and  
identity, as well as identifying information about the  
survey itself. This is set forth as follows.

```
25
(answer
  (surveyName "PBX Satisfaction")
  (surveyDate 12/17/00)
  (surveyVersion "1.0")
30  (questionID 3)
  (type behavioral)
  (time
    (tense current)
    (startDate 12/17/00))
```

```
)
(activity
  (action Contact)
  (queryRegarding Frequency)
5   (indirectObject "NEC/person/salesman")
    (object "NEC/product/PBX/NEAX2000"))
(response
  (type Selection)
  (askingAbout Frequency)
10  (selections (choseOne 0 3 6 9 12 18))
    (primitiveInterval Month)

  (userSelectionRaw 3)
  (userSelection 3)
15  (userSelectionType Month)
    (userID 127)
    (userIDinternal 4208)
    (userIDconfidential "mhcoen@intellistrategies.com: uid
20  0xcf023a8b7")
    (client "NEC/CNG")
)
```

In this way, engine 30 is able to interpret the responses to survey questions using tags. The response  
25 information is analyzed, as described above, to generate graphical displays and recommend follow-up surveys.

## Appendix II

Over-performance and under-performance graphs are components of the report card. The under-performance display is generated according to the process in section 5 1.0 below and the over-performance display is generated according to the process in section 2.0 below.

### 1.0 Under-Performance Display

- 10 For client company b (who is running engine 30) &  
For each competitor company c &  
For each feature f &  
For each user u
- 15 Such that we know:
- (1) (importance of f to u)
  - (2) (satisfaction rating of company b on feature f to user u)
  - 20 (3) (satisfaction rating of company c on feature f to user u)
  - (4) (all involved data is less than 2 months old)

Calculate:

- 25 (1) average and variance of satisfaction for each feature over all competitors c)  
Call these quantities  $avg(f)$  and  $stddev(f)$  respectively
- (2) (average of satisfaction for each feature for company b)
- 30 Call this quantity  $avg(f,b)$

Sort features by importance and proceed through them in decreasing order:

```
5      If (avg(f) - avg(f,b) > stddev (f))  
      Then set rank(f) = (sqrt(importance(f))*(avg(f) -  
      avg(f,c))) - penalty(avg(f), stddev(f)^2)
```

We also subtract a penalty term from the  $rank(f)$  to discount features with high variance either at the moment  
10 (as shown here) or historically.

**Loop:** Consider the  $n$  features with the highest rank, where  $n$  is the number of features to be displayed in the under-performance graph. If any of them are proxies for a  
15 derived attribute, here a feature, and the other proxy attributes are known, calculate the rank for the derived feature and use it instead.

Go to **Loop**.

20 If not, continue.

Generate a chart or graph for each feature and display the features in reverse order by rank.

## 2.0 Over-Performance Display

For client company b (who is running engine 30) &  
For each competitor company c &  
For each feature f &  
5 For each user u

Such that we know:

- (1) (importance of f to u)
- 10 (2) (satisfaction rating of company b on feature f to user u)
- (3) (satisfaction rating of company c on feature f to user u)
- (4) (all involved data is less than 2 months old)

15 Calculate:

- (1) (average and variance of satisfaction for each feature over all competitors c)  
Call these quantities  $avg(f)$  and  $stddev(f)$  respectively
- 20 (2) (average of satisfaction for each feature for company b)  
Call this quantity  $avg(f,b)$

Sort features by importance and proceed through them in

25 increasing order:

If  $(avg(f,b) - avg(f) > stddev(f))$   
Then set  $rank(f) = (sqrt(max - importance(f)) * (avg(f) - avg(f,c))) - penalty(avg(f), stddev(f)^2)$   
30

We also subtract a penalty term from the rank to discount features with high variance either at the moment (as shown

here) or historically. *Max* represents the maximum feature value (i.e., as determined by the source question's scale).

**Loop:** Consider the *n* features with the highest rank. (*n* is  
5 the number of features to be displayed in the  
underperformance graph). If any of the *n* features are  
proxies for a derived attribute (here a feature) and the  
other proxy attributes are known, calculate the rank for  
the derived feature and use it instead.

10

Go to **Loop**.

If not, continue.

Generate chart for each feature and display them in reverse  
15 order by rank.

Appendix III

Surveys by nature are very specific documents. They are written with respect to a particular inquiry, to a specific industry (or entity), to a particular product, offering, or concept, and for an intended audience of respondents. These determine not only the structure of the overall survey but the particular choice of wording in the questions and the structure, wording, and scale of the question answers.

The system (engine 30) has a library of surveys that it can deploy, but instead of containing the actual text of each of their questions, the surveys contain question templates. Each of these templates captures the general language of the question it represents without making any commitment to certain particulars. The system fills in the details to generate an actual question from a template using an internal model of the client who is running the survey that is created during engine 30's configuration for that client. This model includes the client's industry, product lines, pricing, competitors, unique features and offerings, resellers, demographic targets, customer

segmentations, marketing channels, sales forces, sales regions, corporate hierarchy, and retail locations, as well as general industry information, such as expected time frames for product/service use, consumption, and  
5 replacement.

Although generating the question templates requires more effort than simply writing questions directly, it avoids the effort of customizing and modifying every The system survey for each new client.

10 The following are examples of survey questions and the templates that generate them:

1) Purchase frequency:

- 15 a. *How many laptop computers have you purchased in the past 10 years?*  
b. *How many airlines tickets do you buy per year?*

```
(question (variables ${CURRENT_PRODUCT}
${PURCHASE_INTERVAL})
20 (text
    "How many ${CURRENT_PRODUCT} "
    (if ${PURCHASE_INTERVAL} == 12) {"do you buy per
    year"}
    elseif ((mod ${PURCHASE_INTERVAL} 12) == 0)
25 {"have you bought in the past "
    (${PURCHASE_INTERVAL} / 12)
    " years"}
    else {"have you bought in the past
    ${PURCHASE_INTERVAL} months"}
30 )
    "?")
```



)

2) Competitive Position/reliability:

a. Which brand of PBX do you think is most reliable?

- 5                   ☐ NEC  
                  ☐ Nortel  
                  ☐ Lucent  
                  ☐ Williams

10           b. Which type of vehicle do you think is most  
              reliable?

- ☐ Pickup Truck  
                 ☐ SUV  
                 ☐ Station wagon  
15               ☐ Sedan

(question (variables \${CATEGORY\_REFERENCE}  
          \${CURRENT\_PRODUCT}

20                                   \${MANUFACTURERS}))

          (text

                  "Which \${CATEGORY\_REFERENCE} of  
                  \${CURRENT\_PRODUCT} do you think is most reliable?"

25                   )  
                  (scale (selections \${MANUFACTURERS}))

          )

)

30